DOCUMENT RESUME

ED 082 142 CS 000 734

AUTHOR Parnes, Sidney J.; Treffinger, Donald J.

TITLE Development of New Criteria for the Evaluation of Creative Studies Programs.

INSTITUTION State Univ. of New York, Albany. Research Foundation.; State Univ. of New York, Buffalo. Coll. at Buffalo.

SPONS AGENCY Office of Education (DHEW), Washington, D.C.

BUREAU NO BR-2-B-019

PUB DATE Feb 73

GRANT OEG-2-2-2B019

NOTE 46p.

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS College Students; Creative Ability; *Creative Development; *Creative Thinking; *Divergent Thinking; Educational Research; *Problem Solving; *Productive Thinking

IDENTIFIERS *Creativity Training Research

ABSTRACT

The first purposes of this project were to develop a new measure of Creative Problem-Solving (Improving Research Testing) and to develop new scoring criteria for the College Situations Problems. In addition, evidence concerning the reliability and validity of nine new variables derived from these two measures was examined. Adequate levels of inter-problem and inter-rater reliability were obtained for each of the nine variables. Seven of the nine new variables were also correlated positively and significantly (although moderately) with a variety of external criteria commonly associated with creativity, including divergent thinking abilities and several scales representing non-academic accomplishments. Finally, the nine new variables were applied in comparing four experimental groups and a control group in the Creative Studies Project. Significant results, favoring experimental groups, were obtained for three of the five variables derived from the Improving Research Testing measure, but not for the four variables derived from the College Situations Problems. Implications for subsequent research were also identified. (Author/LL)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

ED 082142

DEVELOPMENT OF NEW CRITERIA FOR THE EVALUATION OF CREATIVE STUDIES PROGRAMS

Sidney J. Parnes
State University College at Buffalo
and
Donald J. Treffinger
University of Kansas

Project Number 2B019
Grant Number OEG-2-2-2B019

State University College at Buffalo
Buffalo, New York 14222
and
Research Foundation of State University of New York
Albany, New York 12224

The Research Reported Herein Was
Supported by a Grant from the
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education (Region II)

February, 1973

CS 000 734

TABLE OF CONTENTS

	Page
List of Tables	iv
Abstract	v
*CHAPTER I: Introduction and Statement of Problem	1
Review of Literature	2
Facilitation: Techniques and Programs	2
Limitations of Existing Measures	3
Development of New Measures: Current Research	5
Summary	6
The Creative Studies Project	6
Specific Objectives of the Study	7
*CHAPTER II: Method	8
Criteria for Development of New Measures	8
Development of A New Problem	9
Rationale and Scoring Criteria	9
Workability	10
Importance	11
Development of New Scoring Criteria	11
Sample Situation	12
Fluency	12
Flexibility	12
Originality	14
Use of Structural Analysis	14
Locus of Control	15
Knowledge and Use of CPS Methods	15
Procedures: Investigation of Validity and Reliability	16
Sample	16
Instruments	17
Divergent Production and Transformations Tests	17
Non-Academic Accomplishments	17
The Adjective Check List	17
Statistical Procedures	17
Inter-Rater Reliabilities	18

	Page
Inter-Variable Correlations	18
Correlations with External Criteria	18
Comparison of High and Low Criterion Groups	18
Procedures: Application in Creative Studies Project	18
Summary	19
*CHAPTER III: Results	20
Reliability Tests	20
Inter-Rater Reliability	20
Inter-Problem Reliability	21
Correlations with External Criteria	21
Improving Research Testing	22
College Situations Problems	23
Summary	24
Total Non-Academic Accomplishments	24
Application in the Creative Studies Project	25
Improving Research Testing: Fluency, Part I	25
Improving Research Testing: Fluency, Part II	26
Improving Research Testing: Workability	27
Improving Research Testing: Importance	27
Improving Research Testing: Overall Rating	28
College Situations Problem: Fluency	28
College Situations Problem: Flexibility	29
College Situations Problem: Originality	29
College Situations Problem: Structural Analysis	30
*CHAPTER IV: Discussion and Conclusion	31
Objective One	31
Objective Two	31
Objective Three	31

	Page
Objective Four	32
Objective Five	33
Directions for Future Research	33
Summary and Conclusions	34
References	35
Appendix A: Preliminary Scoring Chart for Improving Research Testing, Part III	39
Appendix B: Means and Standard Deviations For All Project Variables	40

LIST OF TABLES

Table		Page
1.	Discrepancies Between Implications of Knowledge About Creativity and Conventional Testing Procedures	4
2.	College Situations: Frequency Distribution of Responses	14
3.	Inter-Rater Reliability for All New Variables	20
4.	Intercorrelations Among New Variables	21
5.	Correlations of Nine Newly-Developed Variables With 23 Selected External Creativity Criteria	23
6.	One-Way ANOVA Comparisons of High and Low Criterion Groups on The American College Survey (Total Accomplishments) for New Variables	25
7.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison of Fluency, Part I	26
8.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison of Fluency, Part II	26
9.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on Workability	27
10.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on Importance	27
11.	Means, Standard Deviations and ANOVA: Experimental-Control Comparison on Overall Rating	28
12.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on College Situations Problem: Fluency	28
13.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on College Situations Problem: Flexibility	29
14.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on College Situations Problem: Originality	29
15.	Means, Standard Deviations, and ANOVA: Experimental-Control Comparison on College Situations Problem: Structural Analysis	30

ABSTRACT

Researchers and program developers in the area of creative problem-solving must identify criteria in which complex, real-life problems are involved in order to demonstrate the effectiveness of instructional programs which purport to improve creative problem-solving abilities. The first purposes of this project were to develop a new measure of Creative Problem-Solving (Improving Research Testing) and to develop new scoring criteria for the College Situations Problems.

In addition, evidence concerning the reliability and validity of nine new variables derived from these two measures was examined. Adequate levels of inter-problem and inter-rater reliability were obtained for each of the nine variables. Seven of the nine new variables were also correlated positively and significantly (although moderately) with a variety of external criteria commonly associated with creativity, including divergent thinking abilities and several scales representing non-academic accomplishments.

Finally, the nine new variables were applied in comparing four experimental groups and a control group in the Creative Studies Project. Significant results, favoring experimental groups, were obtained for three of the five variables derived from the Improving Research Testing measure, but not for the four variables derived from the College Situations Problem. Implications for subsequent research were also identified.

CHAPTER I
INTRODUCTION AND STATEMENT OF PROBLEM

In the two decades since Guilford's (1950) discussion of the lack of research interest in creativity, activity among educators and psychologists has increased rapidly. One of the most extensively studied problems is whether or not it is possible to nurture creativity through training or instruction. Many methods and techniques have been advanced, and several instructional programs and curricula have been developed. These purport to have, either as their principal goal or at least as a goal of some significance, the facilitation of creativity (Treffinger and Gowan, 1971). Utilizing many of these methods and programs, numerous experimental tests have been conducted, particularly among children and adolescents. The results of many studies suggest that it is possible to facilitate many divergent thinking abilities through deliberate training or instruction.

It seems very important, however, to inquire whether or not such training attempts also have more complex effects. However important the divergent thinking abilities may be as a necessary aspect of the measurement of creativity (Guilford, 1967a, 1971), the complexity of the creative process indicates that many different abilities, and some non-aptitude traits, influence in significant ways the development and expression of creativity. Although the divergent thinking abilities are important criteria in creativity-training research, they should not be considered sufficient criteria (Treffinger, Renzulli, and Feldhusen, 1971).

Even though there is evidence that divergent thinking abilities are correlated with creative productivity, it does not necessarily follow, of course, that training which facilitates divergent thinking will also result in increases in creative behavior (more generally defined). The question is, then, "Will increasing divergent thinking also result in increased effectiveness in more complex creative problem-solving?" Of what use are increased divergent-thinking abilities, we might ask, unless they contribute to a person's ability to solve more quickly or effectively the many complex problems which life constantly presents?

There are also several conceptual and methodological problems relating to the use of many existing measures of divergent thinking, which further suggests that additional criteria are needed for a comprehensive assessment of the effects of a training program.

Critical problems for the researcher or program developer, then, are to identify criteria which represent the complex problem situations with which the subject must deal in his ordinary life experiences (i.e., "real life" or "naturalistic" problems), and to demonstrate that the training has had a facilitating effect on such criteria, as well as on less complex measures (such as specific divergent thinking tasks). This represents, in one sense, a step toward the ultimate criterion for creativity training research: does the training influence the person's behavior in naturalistic problem-solving settings?

Although it might be most advantageous and enlightening to follow a group of trained and control subjects over a long period of time, as they deal with their real problems, such procedures are clearly impractical. In the proposed project, therefore, an attempt will be made to develop and validate new criteria which will meet the requirements of research utilization, and, at the same time, assess outcomes more complex that those assessed by existing measures of creative and divergent thinking. These criteria will then be utilized in the overall evaluation of the Creative Studies Project, currently being conducted at the State University College at Buffalo, New York (Parnes and Noller, 1973).

Review of Literature

In this section, two important areas of concern in the proposed research will be reviewed: first, literature pertaining to techniques and programs for facilitating creative behavior; and second, recent theory and research concerned with problems in the measurement of creative thinking and problem solving.

Facilitation: Techniques and Programs. Wide-spread interest in creativity in the last two decades has resulted in the development of many procedures, methods, and programs for nurturing creativity. Treffinger and Gowan (1971) have listed over forty different approaches and programs. Among these, many have been developed as instructional programs for utilization in school settings. Numerous research studies have been conducted in which their effectiveness has been tested under a variety of conditions.

Torrance and his associates (Torrance and Gupta, 1964; Cunnington and Torrance, 1965; Myers and Torrance, 1965) have developed printed and recorded programs for fostering creative thinking among children and early adolescents. An instructional program for elementary school children, in which several techniques are integrated, has been developed and tested by Davis and Houtman (1968). Covington, Crutchfield, and Davies (1966) developed programed instructional materials for facilitating creative problem-solving among fifth- and sixth-grade pupils; these materials have been utilized in several research investigations (Treffinger and Ripple, 1971). The Purdue Creativity Training Program (Feldhusen, Treffinger, and Bahlke, 1970), is a series of 28 audio-taped programs, each with accompanying printed exercises, to facilitate creative thinking and problem solving among pupils in the upper elementary grades. Evidence for its effectiveness has been presented in several studies, including a large-scale test involving forty-eight fourth-, fifth-, and sixth-grade classes (Feldhusen, Treffinger, and Thomas, 1971). Synectics Educational Systems has also produced a broad range of educational materials, designed to facilitate creative talent among students at many age levels, from childhood to adult (Gordon, 1971).

One of the most extensive programs of research and development, particularly in relation to training programs for high school students, college students and adults, has been the program developed at Buffalo by Osborn (1953) and continued and expanded by Parnes and his associates (1967a). Their efforts have included a wide range of programs and research investigations, including:

(1) establishment and evaluation of a college-level elective in creative problem-solving (Parnes and Meadow 1959; Meadow and Parnes, 1959);
(2) a week long Creative Problem-Solving Institute, for adults in industry, government, and education. This has recently completed its eighteenth annual program (Journal of Creative Behavior, Vol. 5, No. 4, 1971, pp. 281-290.
(3) and, an experimental program with high school pupils (Parnes, 1966).

There is now in progress at the State University College at Buffalo, New York, a Creative Studies Project, which extends this line of research (Parnes and Noller, 1973). Incoming freshman students who applied for participation in the Project's program were randomly selected and assigned to experimental and control groups. Students in the experimental condition participated in a four-semester sequence of courses in creative problem-solving. Assessments of their growth in creative abilities were made, including comparisions of experimental and control subjects on a wide range of cognitive and affective variables (Parnes and Noller, 1973). Longitudinal comparisons are also planned. (Experimental group instruction began in September 1970.)

Research with many of the programs discussed above leads quite clearly to the conclusion that it is possible to effect significant facilitation of divergent-

thinking abilities. Many of the studies cited above, using a variety of training procedures and programs, have shown quite consistently that creative thinking abilities can be nurtured (Meadow and Parnes, 1959; Britton, 1967; Torrance, 1965; Torrance and Myers, 1970; Feldhusen, Treffinger, and Thomas, 1971; Parnes, 1967b; Guilford, 1967). There is also some evidence that the effects of training persist over time (Covington and Crutchfield, 1965; and Parnes and Meadow, 1960), and that instructed Ss are superior to controls in performance on several problem-solving tasks (Olton, 1969; Ripple and Dacey, 1967; Treffinger and Ripple, 1971) and on measures of personality variables and attitudes (Parnes and Meadow, 1959; Treffinger and Ripple, 1969). There is very little empirical evidence, however, concerning the effects of training programs or procedures on more complex, naturalistic or "real-life" problem-solving criteria.

Limitations of Existing Measures. Researchers have been concerned with the development of valid, reliable, useable measures for creativity for many years (e.g., Guilford, 1950, 1967a; Torrance, 1962, 1966; Mednick, 1962), and it is clear that substantial progress has been made in this area (Guilford, 1967b; 1970). However, as Treffinger, Renzulli and Feldhusen (1971) pointed out, there are many problems which still remain unsolved.

The measures of fluency, flexibility, originality, and elaboration which are yielded by several existing measures (Guilford, 1967a; Torrance, 1966) do not provide a comprehensive measure of creative potential. Divergent thinking scores appear to represent a necessary, but not sufficient, component of the assessment of creativity. Guilford (1967a, 1971) has argued explicitly, for example, that (in Structure-of-Intellect terminology) other abilities, including those which involve transformations as products, and several which involve behavioral content, are very likely to be important aspects of creativity. In addition, a number of personality and affective variables are very likely to be important components of creative potential (Dellas and Geier, 1970).

Covington (in press) described many aspects of the process of creative problem-solving, and concluded that measures of divergent thinking, or any attempt to assess creativity through customary psychometric "objective-testing" procedures, would be likely to be inadequate. He contended that, particularly because of the complexity of the creative process, any attempt to assess creativity which relied exclusively on "factorially pure" tests of specific aptitudes would be non-comprehensive. Guilford (1971) also argued that no single aptitude, nor even a set of measures of similar aptitudes, such as divergent thinking abilities, should be expected to correlate very highly with other criteria of creative potential.

Thus, existing measures of creative thinking seem, logically, to be limited in value. Certainly they constitute one important criterion to be considered in creativity training research, but they should not be the sole criteria upon which the researcher or developer relies.

There are also technical and methodological limitations of existing measures, which bear importantly on our use and interpretation of such measures. Existing measures have been criticized and problems identified by Ausubel (1963), Wallach and Kogan (1965), Wallach (1968), Thorndike (1963), Vernon (1964), Harvey (1970), Clark and Mirels (1970), Taft and Rossiter (1966), Guilford (1971), Treffinger, Renzulli, and Feldhusen (1971), Khatena (1971), Tryk (1968) and Covington (in press). In general, the problems which are most frequently discussed involve:

(1) Problems of reliability, especially in relation to the test-retest reliability of existing measures, which is frequently substantially lower than desired over short intervals, and is seldom as high as reliability coefficients for measures of other cognitive abilities (such as IQ).

(2) Problems of <u>useability</u>, which have to do with such concerns as the effects of variations in time limits, testing conditions, administration procedures and directions, as well as issues concerning the objectivity and accuracy of scoring.

(3) Problems concerning the <u>internal structure</u> and <u>construct validity</u> of existing measures, including those deriving from factor analytic investigations, and such issues as the comparability of sub-tests and the combination of scores derived from several sub-[illegible].

(4) Problems of <u>predictive and concurrent validity</u>, with two particular common concerns: (a) that there is insufficient evidence that scores on existing measures either predict [illegible] creative accomplishments or are significantly correlated with other acceptable concurrent criteria of creative behavior; or (b) that scores on such [illegible] commonly intercorrelate better with measures of other aptitudes than among themselves or among other measures of similar aptitudes.

(5) Problems concerning the <u>content validity</u>, or the theoretical and conceptual adequacy of tests in relation to what is known about creative potential from other sources. Covington (in press), for example, has identified several ways in which conventional psychometric procedures seem unsuitable for assessing creative ability, and hence, several ways in which many existing measures are limited; these are summarized in Table 1.

Table 1

Discrepancies Between Implications of Knowledge About Creativity and Conventional Testing Procedures
(Based on Covington, in press)

CREATIVITY IMPLIES:	CONVENTIONAL TESTING:
1. Deep, personal commitment and involvement in problem	1. Artificial, highly contrived situations
2. Complex problems, unique to the person	2. Formalized, simplified, standardized tasks
3. Serious effort; personal consequence involved	3. Impersonal, frequently frivolous tasks
4. A single problem considered, with intense consideration followed by ample time for incubation	4. A large number of discrete items, with relatively short periods of working time
5. A variety of cognitive and affective variables interplay	5. Emphasis on some "pure" single aptitude or ability
6. Emphasis on coordination and management of ideas, feelings, thought processes	6. Emphasis on products
7. Initially: a "mess"; followed by spontaneous organization and manipulation, without external direction to do so.	7. Clear, concise directions: subject knows what to do, how to proceed, the nature of the needed products, etc.

The present state of affairs with respect to the measurement of problem-solving is also complex. Perhaps the most concise evaluation of that area was made by Davis (1966): "chaotic". Many studies have been conducted in which the criterion was a task constructed especially for the study; such criteria are seldom used by other researchers, and frequently lack evidence concerning validity and reliability. In addition, attempts are seldom made to analyze the psychological abilities or factors assessed in a complex problem-solving task. Little attention has been given to any analysis of the performance called for by the problem-solving task in relation to the structure or content of the training program or experimental manipulation. Yet at the same time, research has indicated that there does not appear to be a single unitary problem-solving aptitude (Guilford et al., 1962). Although Davis (1966) proposed a dichotomous classification scheme for problem-solving tasks, there has yet been no empirical verification of the hypothesized dimensions. Finally, since the passing from general practice of introspective analyses of behavior during problem solving, there have been few attempts to distinguish processes which are involved in a complex problem-solving sequence, or to assess the subject's motives, feelings, or reactions during the problem episode. Thus the criterion has frequently been reduced to a simple dichotomy (solution/no solution), which results in the loss of substantial information which might be important in the evaluation of an experimental training program.

Thus, it is clear that many difficulties confront the researcher or developer in the area of creative training. Existing measures purporting to measure creative thinking abilities provide useful psychological information, and are important components of the overall evaluation of the effects of training. But they are also limited in a variety of ways. Existing problem-solving tasks are also limited in many important ways; there does not appear to exist any single instrument or battery of tests for comprehensive assessment of problem-solving abilities.

Development of New Measures: Current Research. As in any rapidly expanding area of research, it is very difficult to ascertain the number of projects now in progress in which new measures of creative thinking and problem solving are being developed and tested. However, in the process of developing the Creative Studies Project, the Principal Investigator, with the assistance of several other researchers in the area of creativity, conducted an extensive search for, and evaluation of, criterion measures. In addition, the Creative Education Foundation, through its publication, the Journal of Creative Behavior, searches widely for appropriate information concerning creativity and its measurement. As a result of these efforts, several projects have been identified which are related to the current problem.

Covington (in press), and his associates at the Berkeley Creativity Project, have attempted, for example, to develop tests of creative problem-solving in a programed instructional format. In these tests, the subject is confronted with a problem situation. As he progresses through the test, it requires him to generate hypotheses about the problem, test them against known and puzzling facts, account for discrepancies and new information, and, eventually, to describe a final solution. Such measures are in an early stage of development, however, and relatively little is known about their validity and reliability. In addition, these tests are being developed for use with fifth- and sixth-grade pupils, rather than with adolescents and adults. They may also be limited by other factors, such as the fact that "clues" provided for the subject are frequently rather obvious analogues to the solution.

Purdue University's "Belmont Project" (Asher et al., 1970) has been working on the development of new measures of complex cognitive functioning. This project, however, is also in an early stage of development, and, in addition, focusses its efforts only on the development of measures for accurate assessment of specific abilities among disadvantaged groups of elementary school children.

Miles (1968) has attempted to develop "real life" criteria for evaluating the effectiveness of training in problem solving. However, the measures developed were highly specific to the training program involved (e.g. creative design students; or Peace Corps Volunteers), and little validity or reliability information has been provided.

In searching for criteria for use in the Creative Studies Project, over 70 possible measures were considered. Alt ough no single criterion measure was identified which was determined to be entirely satisfactory in assessing the overall effects of the training program, two promising criteria were found (even though each was originally designed for a somewhat different purpose).

Goldfried and D'Zurilla (1969) developed a series of college-related problematic situations, in which the subject receives a brief description of a problem that might be experienced by any student on a contemporary American university campus (i.e., problems relating to courses, grades, dormitory relations and rules, relationships with parents and peers, dating, drugs, etc.). Their purposes were to examine subjects' responses to these cases according to criteria for competence or effectiveness of personal behavior, and to assess how effectively students cope with such problems. The structure of the test also seems promising, however, as a measure of creative problem-solving ability.

Coelho, Silber, and Hamburg (1962) reported on the development of a projective instrument, called the Student TAT, to assess coping behavior among college students. Subjects were presented with pictures representing 10 ambiguous college situations. Their ability effectively to solve specific, potentially stressful problems was assessed from their verbal descriptions of the situation and their description of the behavior of the characters. Again, although the authors' purposes were very specifically to compare various groups on the effectiveness with which they cope with problems, there seems to be promise for using such an instrument as an assessment of creative problem-solving ability per se.

Summary. Although divergent thinking measures represent one kind of appropriate criterion for assessing creativity-training, they are limited by several conceptual and technical problems, and should not be considered a sufficient criterion. It was therefore postulated that additional criteria are needed which are valid, reliable, useable, based on adequate psychological theory, and constructed so as to be appropriate to the content of the training program. In addition, such criteria should assess the subject's behavior in a situation which is related to real-life problems, and which is not perceived by the subject as arbitrary or contrived. Two measures, developed for other purposes, seem to have potential value for utilization as criteria for assessing creative problem solving, although neither of these, nor any of several other current attempts at developing new measures, is entirely adequate in its present form for creativity assessment.

The Creative Studies Project

The Creative Studies Project of State University of New York College at Buffalo (Noller and Parnes, 1973) represents one of the most extensive projects involving instruction in creative problem-solving that has ever been conducted. In addition to following the performance of experimental subjects, using a variety of tests and inventories, through four semesters of instruction, the evaluation of the effectiveness of the instruction has been carefully controlled. In addition to utilizing a control group which has also been tested over the entire four semester instructional period, great effort has been made to continue to study the performance of subjects who withdrew from the experimental program after one, two and three semesters. A major focus of the present project, then, was to develop new, complex criteria for assessing the effectiveness of instruction in creative

problem-solving among college undergraduate students.

Specific Objectives of the Study

As part of the general goal of developing new, complex criteria for measuring creative problem-solving abilities, the following specific objectives were formulated:

(1) To develop new scoring criteria in which specific creative abilities are assessed, for problem situations that have already been constructed by Goldfried and D'Zurilla (1969);

(2) To develop a new measure, in which subjects are asked to solve "real-life" problems related to their experiences in the experimental project;

(3) To investigate the interrelationships of the scores obtained from the measures described in Objectives One and Two.

(4) To investigate the validity of the newly-developed variables, indicated by their relationship with other appropriate external criteria of creativity.

(5) To employ the newly-developed variables in the Creative Studies Project in order to evaluate their unique contributions to the assessment of the effects of the experimental program.

CHAPTER II
METHOD

The purposes of this chapter are:

(1) to describe the criteria for developing new measures;
(2) to describe the development of new variables for assessing creative problem-solving;
(3) to describe the development of creative problem-solving scoring criteria for the College Situations problems;
(4) to describe the procedures which were employed for investigating the validity and reliability of the newly-developed variables; and
(5) to describe the procedures for the application of the newly-developed indices of creative problem-solving in the Creative Studies Project.

Criteria for Development of New Measures

The first step in developing new measures for assessing creative problem-solving in the Creative Studies Project was to establish general criteria for the development of measures. Following Miles (1968) and several criticisms of existing measures reviewed in Chapter One, the following criteria were established:

(1) Relevance. The measure must sample from a domain of experiences and problems familiar to all subjects, to enable them to become actively involved in the problem as well as to use previous experiences and training. The performance required by the measure should be similar to other performances commonly involved in problems encountered by the subjects.

(2) Scoring Criteria. The measure must include at least two general scoring dimensions: first, there must be an "effectiveness" dimension (which involves the possibility of actually implementing a solution to the problem presented) and secondly, there must be a "creative" dimension which may involve one or more of the following specific criteria:

(a) fluency - the ability to enumerate many ideas related to the problem;
(b) flexibility - the ability to "shift" readily among several kinds or classes of ideas and solutions;
(c) originality - the ability to produce unusual or uncommon ideas and solutions.

(3) Variety of Solutions. The problems should be "open-ended", so that many different ideas and solutions can be given.

(4) Problem-Solving Time. Every subject should have adequate time to solve the problem(s) presented in the new measure. The problems must be long enough to provide a challenge, but not so long as to lead to "fatigue" when administered as part of a battery of experimental tests.

(5) Experimental Control of Resources. All materials and resources necessary for the development of adequate solutions for the problem should be present and provided by the experimenter in the administration of the new measure.

(6) Reliability. Scores derived from new measures should be significantly and positively intercorrelated, and reliable scoring should be possible with minimum training and time expenditure.

(7) Complexity and Reality. The tasks or problems should represent moderately complex situations, and consequences of solutions should be real and able to be implemented. The problems should not be excessively formalized.

simplified, impersonal or frivolous.

(8) Number of Tasks. Rather than presenting the subject with a large number of discrete items, the problem should be a single problem, or a small set of interrelated problems.

(9) Variety of Skills, Traits, and Abilities. The problem should not emphasize, through content or directions, any single skill, ability, or trait (such as ideational fluency, listing ideas, or finding "wild" ideas).

(10) General Directions. The directions for the problem should not be deliberately confusing or misleading, but there should be opportunities for the subjects to organize and manipulate the task independently.

These criteria were followed in developing a new instrument and in developing new methods of scoring the College Situations Problems.

Development of A New Problem

One of the first activities of the project was to develop a new measure, which was called, "Improving Research Testing". This measure consisted of three separate sections, presented in immediate succession to all subjects.

The three sections of "Improving Research Testing" were:

(1) Ideas for improving testing, in which the subject is asked to generate as many ideas as possible for improving the overall creativity testing operation.

(2) Making desicions, in which the subject is asked to identify as many factors as possible to use in deciding which of the ideas produced in Part 1 are the best.

(3) Making a plan, in which the subject is to apply his criteria to his ideas, and thereby formulate a specific proposal for the improvement of research testing.

Rationale and Scoring Criteria. The Improving Research Testing problem was constructed for use as a criterion measure in the Creative Studies Project for a variety of reasons. First of all, it asks initially for divergent production in a meaningful context. Although the ability to generate many ideas, to produce many different kinds of ideas, or to produce unusual ideas are not unusual approaches to the measurement of creativity, tasks usually employed involve situations which are not readily familiar in the everyday experience of the subject.

In this measure, however, we feel that an attempt has been made to create a test stimulus which is meaningful to every subject (since all have been participants in the research testing for the Project). At the same time, however, it is unlikely that there will be wide variations in the previous experiences of the subjects, a problem frequently encountered in attempts to develop "real-life" or "relevant" problem-solving tasks. Thus, subjects are presented with a relevant task in which their past experience will have negligible influence on their present performance.

The three stages of the problem provide a measure which can also be shown to relate to the instructional program which has been used in the Creative Studies Project. In asking subjects first to generate ideas, then to develop evaluation criteria, and finally to produce a specific plan, there is a correspondence with sequential creative problem-solving. Since experimental subjects have been taught to defer judgment during ideation, to develop specific criteria for judgment when evaluation becomes appropriate, and to use their ideas and evaluation

criteria in a planful, systematic way, they should be more facile than uninstructed students in meeting the demands made by the three phases of this task. We would expect, therefore, that their ideation will be more productive, their criteria more specific and extensive, and their plans more effective and better able to be implemented.

For Part 1, the purpose of which was to provide an opportunity for subjects to produce as many ideas as possible about the improvement of research testing, it was concluded that the most direct and appropriate measure was a simple fluency score.

For Part 2, the purpose of which was to provide an opportunity for the subjects to identify as many evaluation criteria as possible, it was again concluded that a fluency measure should be employed. Since the purpose of Part 2 was not to ask subjects to select any particular criteria or categories, nor to apply the criteria to their previous list of ideas, the proper focus appeared to be the number of criteria produced.

For Part 3, in which subjects were asked to formulate a plan, the selection and development of scoring procedures was much more difficult. In accord with the criteria established for developing the new measures, we sought scores that would represent the workability or effectiveness of the subjects' responses, as well as a dimension of originality in their responses. Given a problem that was very real, with which each of the subjects had had considerable previous experience, the emphasis was on identifying responses which were not just original, but functional as well. Responses to real problems which are only unusual -- perhaps to the point of being bizarre -- would not have represented well the goals of instruction in creative problem-solving.

In Part 3, the subjects were asked to use the ideas and the criteria which were identified in Parts 1 and 2, and thus to develop the best overall plan for improving the research testing operation. The purpose of the problem was to determine whether or not the subjects could work with their lists of ideas, including their novel or original ones, and with their criteria, to formulate an effective course of action.

Accordingly, no fluency, flexibility, or originality scores as such were calculated for Part 3; fluency without explanation, or flexibility and/or originality without implementation, were not sought in Part 3. We concurred at this point with MacKinnon's (1962) observation that the genuinely creative response must be adaptive to reality. Many possible alternative scoring criteria were examined and considered, in an attempt to develop scoring criteria that would take into account factors such as: organization and structure, uniqueness, possibility for implementing the plan, and sensitivity to problems and difficulties. An example of one analysis of Part 3, involving five possible scoring dimensions, with seven levels for each dimension, is presented in Appendix A. Unfortunately, it proved impractical to translate these complex dimensions into scoring criteria which could be applied with reliability. Finally, therefore, two basic scoring dimensions were developed, for which reliable ratings could be obtained. These were named "Workability" and "Importance".

Workability. Ease of applying and probability of success, (1) within the prime experimental objective of realistically evaluating growth, and (2) considering the major criteria of cost, college and faculty policies, effects on experimental subjects, and effects on research staff:

(0) Blank paper, irrelevant, hostile, states that everything is "fine as is" (stated or implied).

(1) Completely vague or impractical (in terms of criteria and prime

objective); may be simply a personal complaint, no spelling out, no sensitivity to criteria expressed or implied.

(2) Detailed paper but completely impractical; or, undetailed but shows awareness to impracticality.

(3) Quite impractical as given, serious shortcomings, little or no spelling out or sensitivity.

(4) Quite impractical even though reasonable amount of detail and/or sensitivity to major shortcomings.

(5) Somewhat impractical as given, minor shortcomings exist without spelling out how to overcome them or without showing sensitivity to flaws (may have detail).

(6) Somewhat impractical; shows sensitivity to minor shortcomings but does nothing about them.

(7) Feasible as explained; no serious shortcoming on criteria; or, most apparent shortcomings, if any, reasonably covered by sensitivity and spelling out.

(8) Workable as explained, 5 or more ideas presented, or more than one level of spelling out, sensitivity to new challenges implied or expressed.

(9) Very workable plan, 5 or more ideas expressed, or more than one level of spelling out, sensitivity expressed (sees problems that could arise and solves them).

Importance. In relation to accomplishing overall objective of realistically evaluating growth:

(0) Blank paper, irrelevant, hostile, states that everything is "fine as is" (stated or implied).

(1) Superficial, 1 or 2 trivial ideas, no clear recognition of problem, personal complaint.

(2) No clear recognition of real overall problem. Mostly minor ideas. Complains for self and others (general complaints).

(3) Definition of problem implied; shows some awareness of problem evolving; gives only minor ideas; shows more direction.

(4) Sees at least 1 important aspect of problem -- i.e., administration of tests, the tests themselves, rationale, motivation, scheduling, atmosphere, people involved.

A third score yielded by Part 3 of Improving Research Testing was an overall rating of the subject's research plan. All subjects' responses to Part 3 were read and rated independently by a counseling psychologist and two researchers, all having had extensive experience in creativity research. Each rater was asked to separate the responses into six categories, from poorest to best in quality; each protocol was then assigned a rating from a low of 1 to a high of 6, according to the category-group in which it had been classified. The overall rating score for each protocal was determined by dropping the most extreme of the three individual's ratings and computing the mean of the remaining two ratings.

Thus, five scores or variables were derived from the Improving Research Testing questions:

(1) Fluency - Part 1
(2) Fluency - Part 2
(3) Workability - Part 3
(4) Importance - Part 3
(5) Overall Rating - Part 3

Development of New Scoring Criteria

The second phase of the project involved the development of new scoring criteria for the College Situations Problems (Goldfried and D'Zurilla, 1969). Although

this instrument was developed originally for assessment of psychological adjustment among college students, it seemed well-suited for adaptation as an indicator of creative problem-solving among the subjects in the Creative Studies Project. The problems included were relevant to the life experiences of college subjects. In addition, the problems were open-ended, and could be answered in controlled time periods without special facilities or resources. If reliable scoring criteria for creativity could be established, it seemed that the College Situations Problems represented a useful criterion for assessing the effects of instruction in creative problem-solving among college students.

A sample problem follows:

"As you read the situation, we would like you to imagine that you are now in this situation. When you have the situation clearly in mind, think of how you are most likely to react in such a situation. Then in the space below the situation, write down your total reaction in specific detail."

SAMPLE SITUATION:

"It is about a month after the start of classes during your first semester, and several important examinations have been scheduled for the same week. The examination for your most difficult course has been scheduled for late Wednesday afternoon.

"You are having breakfast on Wednesday morning, the day of your most difficult exam. You feel that you are inadequately prepared, and your full schedule of classes for Wednesday does not allow time for further studying before the exam."

Four variables were eventually derived. These were identified as: Fluency, Flexibility, Originality, and Structural Analysis. Two additional scoring dimensions (Locus of Control and Knowledge and Application of CPS Methods) were originally constructed, but were dropped from the final scoring and analysis because there were so few stable non-zero scores that no usable information appeared to be obtained. Descriptions of each of the scoring dimensions follows:

I. Fluency. This score represents the total number of ideas that the subject produces in response to the problem. "Ideas" include all things which the person says he will do. Award one point for each idea given, as long as some specific action or behavior is described. This also applies to specific actions which are incorporated into more complex responses. That is, an idea should be counted every time the student actually describes something he could do. Include the response, "do nothing about the problem", if the person actually says that this would be his course of action in dealing with the problem. The FLUENCY score is the total number of ideas given.

II. Flexibility. This score represents the subject's ability to see different kinds of possible solutions. . . to see alternative ways of solving the problem.

We shall assess this by classifying the kinds of ideas the subject produces, using the categories listed below. For every idea which is awarded a point for FLUENCY, one or more of these categories will apply. Try to select the category which best describes the major part of the idea being rated, although, if it is really clear, you may decide that one idea fits in more than one category.

So, the person will have category numbers for each idea that was awarded points for fluency. But we cannot just add up the number of points as we did for fluency, for we are concerned here with the number of different categories. Thus, you should go over the list of categories for all the subject's responses, and count a point for each category the first time it appears. Subsequent uses of the same category receive no points. The total flexibility score is therefore the

number of different categories used by the subject.

Most errors in scoring Flexibility will come from two sources: (a) missed categories; and (b) missed student responses. To solve these problems, be certain to study carefully the list of categories, so that you can classify every idea quickly and accurately. Read each paper closely, and remember that any response which has been given a point for fluency must also fit into one or more categories. If it fits into a category which the subject has not previously used, it gets a point for Flexibility.

The Flexibility Scoring Categories are:

(1) Self-Improvement or Change. A solution in which the primary factor is increasing the student's own ability to do something (to think, to feel, or to act). The idea clearly involves self-betterment, and usually describes some cognitive or affective change in the person (I could better myself.... in some way). This category includes direct mention of incentives to motivate self.

(2) Peers. Solutions or ideas which primarily involve peers as the means of solution (example: I'd ask my friends to help me with my homework, etc.... I'd get a new roommate...).

(3) Parents or Guardians as the primary source or factor.

(4) School Advisors or Counsellors as primary source of help or advice.

(5) Counsellors Outside School. Clergy, medical, legal, psychological assistance.

(6) Increasing Tangible Resources. Solve the problem by getting more money, credit, or through possession (attainment) of new products. (Include getting a job, getting a loan, etc.) These responses involve adding some new resources or things to the existing environment, not just better use or modification of what's already there. (Better use of self= category 1; better use of environment or modifications of environment= category 8 or 9) (This category was subsequently combined with category 9).

(7) Group Processes. Solutions which principally involve improvements in interpersonal relations. (Not just getting a friend; emphasis here is on social groups rather than individuals).

(8) Physical Environment. Changes in the structure of one's physical setting or environment (car, dorm, campus, etc.)-- but not just adding new things (see Category 6). Alteration of one's natural, physical surroundings.

(9) Effective Use of Resources. Solutions which involve the natural and physical environment as it is, but stress more effective utilization of resources. (Making what I have work more effectively). Includes time and sleep.

(10) Fantasy. Obvious fantasy (make a money tree; find a long-lost millionaire relative, etc.)

(11) Redefinition. Responses in which the person solves the problem by defining it in a different way. (Make the "problem" go away by looking at it in some new light). (Subsequently combined with category 13.)

(12) Rest and Relaxation. Solutions in which the principal act is avoiding the problem, taking one's mind off it, doing something else, etc.

(13) Do Nothing. (Combined with Category 11).

(14) Study Schedule and Pace. Solutions involving establishing a schedule or pattern for study (when? whether to cram?)

(15) Study Techniques and Aids. Emphasis in the solution on defining better ways or methods for study.

(16) Emotional Release. Action taken as an integral part of the solution. The purpose is some emotional release (e.g., go scream out the window and then return to books.)

(17) Non-Academic Reward.

(18) Anticipation of Future Action.

(19) Cheating.

(20) Avoidance of Social Contact or Interaction.

III. Originality. Each response in the solution was tabulated, and frequencies were established for "key" actions in each response. Originality was computed by weighting the frequency distribution, following the procedure described by Wilson, Guilford, and Christensen (1962).

The distribution of responses and weights is shown in Table 2:

Table 2

College Situations: Frequency Distribution of Responses

RATING	NO. of RESPONSES	FREQUENCIES INCLUDED
3	138	1
2	132	2-3-4
1	119	5 through 18
0	108	19 and above
	497	

Thus, there were (1) 138 responses, each of which were given by only one subject. Each such response was assigned a score of three; (2) responses given by two, three, or four subjects totaled 132, and each received a score of two; (3) responses given by five or more subjects (but not by more than 18 subjects) totaled 119, and each received a score of one; and (4) 108 responses were given by more than 18 subjects, and each of these received a zero score.

After the scoring weights were computed, each subject received an originality score; this represented the sum of the weights assigned for individual responses included in the subject's total solution to the problem.

IV. Use of Structural Analysis. This score was derived by reading the student's entire response to the problem, and then assigning it a rating as follows:

3 points -- The student has a very clear, complex plan, and states explicitly a sequence of several possible (successive or simultaneous) actions that would be included. He then takes each of these ideas, analyzes it, and proposes

several possible courses that might be taken.

2 points -- The student has a clear idea about what he would do, which may include more than one possible action, but it is not clearly organized. His ideas are presented for several courses of action, but not in an explicitly stated sequence.

1 point -- The student merely gives one or more ideas which specify particular actions he might take. There is no evidence of plan or organization to his ideas.

0 points -- No ideas, or completely irrelevant response.

The two scoring dimensions which were subsequently dropped from all analyses were Locus of Control and Knowledge and Application of CPS Methods. The Locus of Control scores turned out to be almost always a "2"; the Knowledge and Application scores could not be derived with satisfactory reliability by scorers without extensive knowledge of the instructional program. The scoring instructions that were attempted were as follows:

V. Locus of Control. Read the entire response given by the student. Ask whether the student feels that the problem is one that can be solved by deliberate efforts on his part, or whether what happens will merely be decided by chance or fate or luck. Award points as follows:

2 points -- Student sees possible solution, and recognizes that he can do something about it that will effect a satisfactory solution.

1 point -- Student sees that problem can be solved but relies exclusively or primarily on others to tell him what to do to solve it. He's not really in command.

0 points -- Student does not feel that there is any course of action that he can take; he feels it is entirely up to others, or to luck, fate, or chance.

VI. Knowledge and Use of CPS Methods. Give 1 point for anything the student says which fits into any of the following areas. (Note that in three categories, you may give more than 1 point-- 1 for each time the category is found. For all others, give a maximum of 1 point per area.)

CPS-METHODS	KNOWLEDGE	USE
FACT-FINDING	Student says he would try to get more information. (1 point)	Student describes a way to get more information which he would use. (1 point)
PROBLEM-FINDING	Student says he would try to state a definition of the problem. (1 point)	Student actually does offer a definition of the "real problem". (1 point)
IDEA-FINDING	Student says he would try to get new ideas, or that he would brainstorm. (1 point)	Student actually does generate a whole list of ideas. (1 point)

CPS-METHODS	KNOWLEDGE	USE
SOLUTION-FINDING	Student says that he would develop criteria for judging his solution, or that he would use certain criteria. (1 point)	Student actually names the criteria or actually uses them to evaluate his ideas. (1 point)
ACCEPTANCE-FINDING	Student says he would try to find ways of making certain his solution would be accepted and used. (1 point)	Student actually describes details for getting his plan accepted. (1 point for each) --- Student names new problems which would develop. (1 point for each) --- Student gives ways of solving these new problems, too. (1 point for each)

Procedures: Investigation of Validity and Reliability

For each of the five variables derived from the Improving Research Testing measure and of the four variables developed for the College Situations Problems, several tests of validity and reliability were conducted. These included:

(1) Inter-score correlations
(2) Inter-rater reliability
(3) Correlation with selected cognitive and affective variables related to creativity, including:
  (a) measures of divergent production and transformations, from the Structure-of-Intellect Model (Guilford, 1967);
  (b) measures of non-academic accomplishments, using the inventory developed by Richards, Holland and Lutz (1966);
  (c) selected items from the Adjective Check List (Gough and Heilbrun, 1965).
(4) Comparison of high and low scorers (total scores) on the inventory of non-academic accomplishments.

The purpose of these tests were to provide preliminary evidence concerning the validity and reliability of the newly-derived measures prior to application of the new variables in the Creative Studies Project.

Sample

The sample for this project consisted of 158 subjects who were participants in the Creative Studies Project (Parnes and Noller, 1973).

The total research sample for the project was randomly selected from over 350 applicants for the Creative Studies program at State University College (Buffalo). These applicants represent approximately one third of the total incoming freshmen group to whom the program was offered. From the total body of applicants, 150 were randomly placed into the experimental group -- organized into six class sections -- to receive the first semester of Creative Studies courses; an equivalent number were randomly assigned to the control group (which received no courses in Creative Studies until the conclusion of the two-year project). All testing of experimental and control subjects was done simultaneously, in late afternoons,

evenings, weekends, or other times when no classes were in session for the subjects.

Instruments

A variety of instruments, related to cognitive and affective components of creative talent, were administered to all subjects during the Creative Studies Project. Scores on these tests provided opportunities to investigate the validity of the newly-derived variables. The additional instruments thus employed were:

(1) Divergent Production and Transformations Tests. Ten measures derived from the Structure-of-Intellect Model, representing the divergent production operation or the transformation products, were administered. These instruments are considered by Guilford (1967), (Guilford and Hoepfner, 1971) to assess cognitive abilities which are positively related to creative talent. Evidence for the validity and reliability for the following tests, which were employed in this project, is summarized ard reviewed by Guilford (1967) and Guilford and Hoepfner (1971):

(a) Alternate Letter Groups (divergent production of figural classes).
(b) Multiple Social Problems (divergent production of behavioral implications).
(c) Utility Test: Ideational Fluency (divergent production of semantic units).
(d) Utility Test: Spontaneous Flexibility (divergent production of semantic classes).
(e) Multiple Behavioral Grouping (divergent production of behavioral classes).
(f) Varied Emotional Relations (divergent production of behavioral relations).
(g) Insight Problems (divergent production of figural transformations).
(h) Verbal-Picture Translations (cognition of semantic transformations).
(i) Homonyms (memory of semantic transformations).
(j) Jumbled Words (evaluation of symbolic transformations).

(2) Non-Academic Accomplishments. The American College Survey (Richards, Holland, and Lutz, 1966) is a self-report instrument, which asks the subject to report his accomplishments in each of twelve areas: Leadership, Social Participation, Art, Social Service, Science, Business, Humanistic-Cultural Activities, Religious Service, Music, Writing, Social Science, and Speech-Drama.

There is evidence for the inventory's validity (Richards, Holland, and Lutz, 1966), and it seems clear that student accomplishments in the variety of areas included by the twelve categories would be an adequate external criterion of creative expression among college students (cf., Wallach and Wing, 1969).

(3) The Adjective Check List. The Adjective Check List (Gough and Heilbrun, 1965) consists of 300 adjectives commonly used to describe the attributes of a person. These adjectives encompass a wide variety of human traits and behaviors, and 24 separate scales are commonly derived. For the purposes of this project, however, attention was focused upon a set of 22 items. These were selected in advance by the instructors of the creative problem-solving courses because of their particular relevance to creative ability as conceptualized in the instructional program. The newly-developed measures from this study were correlated with the scores derived from this special scale.

Statistical Procedures

After the nine scoring dimensions had been developed for Improving Research Testing and the College Situations Problems, several statistical tests were conducted. For all statistical tests, the .05 level of significance was accepted as satisfactory

for the rejection of the null hypothesis.

(1) Inter-Rater Reliabilities. The first test was to establish satisfactory inter-rater reliabilities for all measures. Raters were trained by the investigators, and were given samples of hypothetical responses until they were familiar with scoring criteria. For each new variable, a sample of 20 tests was rescored, so that inter-rater reliabilities were obtained for all tests and raters. If satisfactory reliability was unable to be attained (as in the complex matrix mentioned earlier for scoring Part III of Improving Research Testing and for Knowledge and Application of CPS Methods for the College Situations Problems), the scoring dimension was dropped from further consideration in the project. We hope that we shall be able to continue revision of these scoring dimensions for possible utilization in future projects. In addition to checking the correlations among the raters' scores, it was further established that there were no significant differences between the scorers' means for each scoring dimension.

(2) Inter-Variable Correlations. The intercorrelations among the five variables derived from the Improving Research Testing measure, and of the four variables derived from the College Situations Problems, were also computed and tested for significance.

(3) Correlations with External Criteria. Next, correlations were computed between each of the nine newly-developed variables and each of the external validity criteria (divergent production and transformation abilities, non-academic achievement, and the special scale of selected items from the Adjective Check List). These correlations were tested for significance.

(4) Comparison of High and Low Criterion Groups on Total Non-Academic Achievement. The final test of construct validity involved a comparison, on each of the nine new variables, of the subjects who scored highest and lowest within the entire sample in total number of non-academic accomplishments. Subjects who scored one standard deviation or more above the mean on total non-academic accomplishments (N=20) constituted the "high" group. Subjects who scored one standard deviation or more below the mean constituted the "low" group (N=19). These groups were then compared on each of the nine new variables separately, using one-way ANOVA.

Procedures: Application in Creative Studies Project

Finally, after the completion of the tests of reliability and validity for each of the nine new variables, each of the newly-developed variables was utilized as a dependent variable for comparing experimental and control groups in the Creative Studies Project. This involved a comparison among five groups:

(1) those experimental subjects having had four semesters of instruction (N=33).
(2) experimental subjects having had three semesters of instruction (N=17).
(3) experimental subjects having had two semesters of instruction (N=20).
(4) experimental subjects having had one semester of instruction (N=31)
(5) control subjects, having received no instruction in creative problem-solving (N=57).

These comparisons, separately for each of the nine new variables, were made using one-way ANOVA, with appropriate post-hoc comparisons.

Summary

In this chapter, the development of the measure, "Improving Research Testing" and the new scoring criteria for the College Situations Problems were described. Procedures for investigating the validity and reliability of nine new variables, and for their application in the Creative Studies Project, were also described.

CHAPTER III
RESULTS

In this chapter, the results of the statistical analyses described in Chapter II will be presented. For each of the nine new variables, the results will be presented for: reliability tests, validation tests, and application in the Creative Studies Project. 1

Reliability Tests

The reliability of the nine new variables was considered in two ways: inter-rater reliability, and inter-problem reliability. The Project testing schedule did not make possible collection of data for determination of test-retest or alternate forms reliabilities.

Inter-Rater Reliability

Reliability of raters was checked through applicable correlation coefficients between the scores of independent raters. As indicated by Table 3, the reliabilities were acceptable, with only two being below the .70's.

Table 3

Inter-Rater Reliability for All New Variables

VARIABLE	$\bar{X}$ SCORER 1	$\bar{X}$ SCORER 2	CORRELATION
A. Improving Research Testing			
1. Fluency Part I	6.7	6.7	.99**
2. Fluency Part II	5.6	5.6	.95**
3. Workability	4.4	4.9	.60**
4. Importance	1.8	1.8	.77**
B. College Situations			
1. Fluency	3.4	3.3	.93**
2. Flexibility	2.7	2.5	.82**
3. Originality	4.7	4.6	.94**
4. Structural Analysis	1.8	1.7	.74**

C. Overall Ratings: Improving Research Testing, Part III

Intercorrelations Among 3 Raters

Rater	1	2	3
1	1.00	.65**	.54**
2		1.00	.53**
3			1.00

** $r > 0$, $p < .01$

1. Means and standard deviations for all experimental variables are presented in Appendix B.

Inter-Problem Reliability

The correlations among the five variables derived from Improving Research Testing and among the four scores derived from the College Situations Problem are presented in Table 4, as well as the correlations between both sets of variables.

Table 4

Intercorrelations Among New Variables[1]

	Improving Research Testing					College Situations			
	(1) Flu-I	(2) Flu-II	(3) Work.	(4) Impt.	(5) Rating	(6) Flu.	(7) Flex.	(8) Orig.	(9) Struct.
(1)	100	70**	13	08	43**	26**	00	19*	15
(2)		100	16	20*	41**	24**	09	10	15
(3)			100	25**	37**	01	-01	-06	01
(4)				100	34**	02	08	02	04
(5)					100	07	01	00	09
(6)						100	55**	55**	42**
(7)							100	48**	41**
(8)								100	25**
(9)									100

1. All decimal points omitted.

* = $p < .05$? r reliably > zero, 156 df.
** = $p < .01$)

The Fluency score from Part I of Improving Research Testing was significantly and positively correlated with the Fluency Score from Part II, with the Overall Rating from Part III, and with the Fluency and Originality scores for the College Situation Problem.

The Fluency Score from Part II of Improving Research Testing was significantly and positively correlated with: Fluency Part I, Importance, and Overall Rating for Part III of Improving Research Testing, and with the College Situations Problem Fluency score.

The Improving Research Testing Workability score was significantly and positively correlated only with the Importance Score and the Overall Rating.

The Improving Research Testing Importance score was correlated positively and significantly with each of the other Improving Research Testing scores except the Fluency Score for Part I.

The Fluency, Flexibility, Originality and Structural Analysis scores for the College Situations Problem were all positively and significantly intercorrelated.

Correlations with External Criteria

Summarized in Table 5 are the correlations of each of the nine new variables with each of 23 external criteria (including 10 selected measures of divergent production or transformation abilities from Guilford's Structure-of-Intellect, 12 scales representing non-academic accomplishments in a variety of areas, and a scale

representing selected items from the Adjective Check List).

Improving Research Testing

The Fluency scores from Part I of the Improving Research Testing measure were significantly and positively correlated with five divergent-production abilities, with three non-academic achievement scales (Social Participation, Humanistic/Cultural, and Writing), and with the selected scale of Adjective Check List items. The Fluency scores from Part II were significantly and positively correlated with four of the divergent production abilities and with three non-academic accomplishment scales (Business, Humanistic/Cultural, and Writing). The Workability score was correlated significantly only with the Social Participation accomplishments scale, and negatively for that scale. The Importance score was positively and significantly correlated with only one measure (cognition of semantic transformations). The Overall Rating was correlated significantly and positively with four divergent production abilities and one transformation ability. It was also significantly negatively correlated with the Artistic accomplishments scale.

(Continued after Table 5)

Table 5

Correlations of Nine Newly-Developed Variables
With 23 Selected External Creativity Criteria

Criterion	New Variable[1] Improving Research Testing Flu-I	Flu-II	Work-ability	Imptce	Rating	College Situations Flu	Flex	Orig	Struct
A. Structure of Intellect[2]									
Alt. Letter Groups (DFC)	09	14	03	-03	00	27**	21**	16*	14
Mult. Social Prob. (DBI)	29**	36**	15	12	28**	22**	12	19*	08
Utilities: Fluency (DMU)	41**	44**	13	-02	23**	29**	12	20*	06
Utilities: Flex. (DMC)	35**	44**	06	-01	16*	21**	10	17*	07
Mult. Behav. Group. (DPC)	18*	15	-03	-12	09	17*	14	09	05
Varied Emot. Rel. (DBR)	29**	33**	00	08	17*	14	03	01	00
Insight Probs. (DFT)	04	02	-05	00	00	02	-02	-05	00
Verbal-Pict. Transl. (CMT)	12	15	03	21**	11	01	02	-03	09
Homonyms (MMT)	-06	00	-04	-01	00	07	09	19*	06
Jumbled Words (EST)	-02	01	11	04	18*	09	05	03	07
B. Non-Academic Accomplishments									
Leadership	11	11	-02	02	07	12	06	05	15
Social Participation	17*	09	-19*	-06	00	19*	02	13	20*
Artistic	00	03	-11	-14	-17*	08	00	-03	-05
Social Service	01	07	-14	13	-01	20*	04	11	23**
Scientific	06	01	-05	05	04	01	-03	05	08
Business	12	16*	01	-09	04	09	-10	04	08
Humanistic/Cultural	29**	22**	-14	00	14	13	00	13	15
Religious	06	13	-07	-13	-10	08	11	05	09
Musical	-07	01	08	-07	-04	02	00	-05	-05
Writing	21**	19*	-07	-03	04	13	07	12	18*
Social Science	15	05	-07	-05	-02	07	-06	12	16*
Speech and Drama	03	-09	-03	-06	-03	-08	-04	04	-06
C. Selected ACL Items	16*	13	-16	-03	-04	-01	-10	01	-01

1. All decimal points omitted

* = $p < .05$ ** = $p < .01$

2. Structure-of-Intellect classification in parentheses.

College Situations Problems

For this measure, the Fluency scores were significantly and positively correlated with five divergent production abilities, and with the Social Participation and Social Service accomplishments scales. The Flexibility scores were significantly and positively correlated only with a measure of divergent production of figural classes. The Originality scores were significantly and positively correlated with four divergent production abilities and with memory of semantic transformations.

The Structural Analysis scores were significantly and positively correlated with four non-academic accomplishment scales (Social Participation, Social Service, Writing, and Social Science).

Summary

Although each of the nine new variables was significantly correlated with at least one of the external creativity criteria, only one of those variables was correlated with a broad range of cognitive, personality, and behavioral criteria (Improving Research Testing-Fluency, Part I). Very few of the new variables seemed to be significantly correlated with transformation abilities, accomplishments in the arts, or the selected set of items from the Adjective Check List. In addition, even when correlations were statistically significant, all were of low to moderate magnitude.

Total Non-Academic Accomplishments

The final criterion for investigating the construct validity of the nine new variables involved the total reported non-academic accomplishments of the subjects. Based on the total number of accomplishments reported in all twelve non-academic scales, high and low groups were identified. The 20 subjects in the high group (with 20 or more reported non-academic accomplishments) and the 19 subjects in the low group (with four or fewer reported non-academic accomplishments)were then compared, using one-way ANOVA, on each of the nine new variables. The results of these comparisons are summarized in Table 6.

The high and low groups did not differ significantly on four of the five variables derived from the Improving Research Testing measure. For the Workability variable, the mean for subjects in the low group (4.72) was significantly greater than the mean for subjects in the high group (2.80).

For the College Situations Problem, subjects in the high non-academic accomplishments group had significantly greater means than subjects in the low group on three of the four new variables: Fluency (4.60 vs. 3.00), Originality (6.70 vs. 4.00) and Structural Analysis (2.25 vs. 1.72).

(Continued after Table 6)

Table 6

One-Way ANOVA Comparisons of High and Low Criterion Groups on The American College Survey (Total Accomplishments) for New Variables.

Criterion Variable	High Group (N=20)[1] $\bar{X}$	S.D.	Low Group (N=19)[2] $\bar{X}$	S.D.	F(p)
Improving Res. Testing					
Fluency - Part I	8.95	5.77	5.88	4.12	3.46 (n.s.)
Fluency - Part II	7.10	4.16	5.44	2.95	1.95 (n.s.)
Workability	2.80	2.39	4.72	2.82	5.14 (<.05)
Importance	1.95	0.75	2.16	0.70	<1 (n.s.)
Combined Ratings	2.65	1.59	2.63	1.52	<1 (n.s.)
College Situations Problem					
Fluency	4.60	1.72	3.00	1.28	10.29 (<.01)
Flexibility	3.10	1.07	2.55	0.85	2.95 (n.s.)
Originality	6.70	2.99	4.00	2.40	9.27 (<.01)
Structural Analysis	2.25	0.55	1.72	0.66	7.11 (<.02)

1. Accomplishments checked ≥ 20 (1 S.D. or more above the mean)

2. Accomplishments checked ≤ 4 (1 S.D. or more below the mean)

Application in the Creative Studies Project

Each of the nine newly-developed variables was utilized as a criterion for comparing experimental and control groups in the Creative Studies Project. There were five comparison groups: experimental groups having had four, three, two, and one semester(s) of instruction respectively, and the controls, who had not participated in any instruction in creative problem-solving. For each analysis, these five comparison groups are identified as groups one through five, in the order named in the previous sentence.

For each analysis, the sample sizes were:

Group 1 (four semester experimentals) - 33
Group 2 (three semester experimentals) - 17
Group 3 (two semester experimentals) - 20
Group 4 (one semester experimentals) - 31
Group 5 (controls) - 57

Improving Research Testing: Fluency, Part I

For the Part I Fluency variable, there was a significant difference ($p < .01$) among the five groups. Post-hoc comparisons using the Newman-Keuls procedure (Winer, 1971) revealed that the 4 semester experimental group's mean was significantly greater than all other groups' means, and that all experimental group means were significantly greater than the control mean. Table 7 reports the means and standard deviations for each group and summarizes the ANOVA for the five groups.

Table 7

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison of Fluency, Part I

Means and Stardard Deviations

Group	1	2	3	4	5
Mean	10.70	8.65	8.40	5.48	4.72
Standard Deviation	4.60	5.37	4.24	3.23	2.72

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	904.09	4	226.02	15.64	<.01
Within Groups	2210.90	153	14.45		
Total	3114.99	157			

Improving Research Testing: Fluency, Part II

For the Part II Fluency variable, there was a significant difference ($p < .01$) among the five groups. Post-hoc comparisons using the Newman-Keuls procedure (Winer, 1971) revealed that all experimental group means were significantly greater than the control mean, that the 4-semester experimental group mean was significantly greater than all other group means, and that the two-semester group's mean was significantly greater than the one-semester group's mean. Table 8 reports the means and standard deviations for all five groups and summarizes the ANOVA comparison.

Table 8

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison of Fluency, Part II

Means and Standard Deviations

Group	1	2	3	4	5
Mean	8.70	6.29	7.00	5.29	3.86
Standard Deviation	2.56	3.53	3.95	2.75	1.74

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	532.54	4	133.14	18.52	<.01
Within Groups	1099.76	153	7.19		
Total	1632.30	157			

Improving Research Testing: Workability

For the Workability scores, there were no significant differences among the five groups. Table 9 reports the means and standard deviations for each group, and summarizes the ANOVA for the five groups.

Table 9

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on Workability

Means and Standard Deviations

Group	1	2	3	4	5
Mean	4.18	2.94	3.20	3.10	3.81
Standard Deviation	2.52	2.56	3.04	2.77	2.56

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	31.9515	4	7.9879	1.13	n.s.
Within Groups	1080.6370	153	7.0630		
Total	1112.5885	157			

Improving Research Testing: Importance

For the Importance variable, there were no significant differences among the five groups. Table 10 reports the means and standard deviations for each group and summarizes the ANOVA for the five groups.

Table 10

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on Importance

Means and Standard Deviations

Group	1	2	3	4	5
Mean	2.21	1.76	1.85	2.03	1.95
Standard Deviation	0.65	0.66	0.75	0.75	0.58

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	3.0409	4	0.7602	1.74	n.s.
Within Groups	66.9338	153	0.4375		
Total	69.9747	157			

Improving Research Testing: Overall Rating

For the Overall Rating score, the four semester experimental group's mean was significantly greater than the other four means. The three and two semester groups did not differ significantly from each other, but they were each significantly greater than the mean for either the one semester group or the controls. The one semester group did not differ significantly from the Control group. Table 11 reports the means and standard deviations for each group and summarizes the ANOVA for the five groups.

Table 11

Means, Standard Deviations and ANOVA:
Experimental-Control Comparison on Overall Rating

Means and Standard Deviations

Group	1	2	3	4	5
Mean	34.55	26.18	26.75	24.36	24.56
Standard Deviation	10.63	18.67	16.17	14.36	12.59

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	2446.95	4	611.738	3.20	.05
Within Groups	29209.53	153	190.912		
Total	31656.48	157			

College Situations Problem: Fluency

For the College Situations Problem, Fluency variable, there were no significant differences among the five groups. Table 12 reports the mean and standard deviation for each group and summarizes the ANOVA for the five groups.

Table 12

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on College Situations Problem: Fluency

Means and Standard Deviations

Group	1	2	3	4	5
Mean	4.09	3.47	3.85	3.32	3.23
Standard Deviation	1.99	1.28	2.43	1.51	1.86

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	19.0199	4	4.755	1.377	n.s.
Within Groups	528.3218	153	3.453		
Total	547.3417	157			

College Situations Problem: Flexibility

For the College Situations Problem, Flexibility variable, there were no significant differences among the five groups. Table 13 reports the mean and standard deviation for each group and summarizes the ANOVA for the five groups.

Table 13

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on College Situations Problem: Flexibility

Means and Standard Deviations

Group	1	2	3	4	5
Mean	2.76	2.76	2.50	2.68	2.65
Standard Deviation	1.32	1.09	0.95	0.98	1.14

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	1.01	4	0.2525	<1	n.s.
Within Groups	193.87	153	1.2672		
Total	194.88	157			

College Situations Problem: Originality

For the College Situations Problem, Originality variable, there were no significant differences among the five groups. Table 14 reports the mean and standard deviation for each group and summarizes the ANOVA for the five groups.

Table 14

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on College Situations Problem: Originality

Means and Standard Deviations

Group	1	2	3	4	5
Mean	5.45	5.00	4.10	4.68	4.75
Standard Deviation	3.83	2.94	2.67	3.44	2.98

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	25.0687	4	6.2672	<1	n.s.
Within Groups	1595.3172	153	10.4269		
Total	1620.3859	157			

College Situations Problem: Structural Analysis

For the Structural Analysis variable, there were no significant differences among the five groups. Table 15 reports the mean and standard deviation for each of the five groups and summarizes the ANOVA for the five groups.

Table 15

Means, Standard Deviations, and ANOVA:
Experimental-Control Comparison on College Situations Problem: Structural Analysis

Means and Standard Deviations

Group	1	2	3	4	5
Mean	2.15	1.94	1.70	1.81	1.86
Standard Deviation	0.57	0.66	0.73	0.54	0.69

ANOVA

Source	S.S.	df	M.S.	F	P
Between Groups	3.2802	4	0.8201	1.99	n.s.
Within Groups	63.0995	153	0.4124		
Total	66.3797	157			

CHAPTER IV

DISCUSSION AND CONCLUSIONS

The specific objectives which were formulated for this project were:

(1) To develop new scoring criteria in which specific creative abilities are assessed, for problem situations that have already been constructed by Goldfried and D'Zurilla (1969);

(2) To develop a new measure, in which subjects are asked to solve "real-life" problems related to their experiences in the experimental project;

(3) To investigate the interrelationships of the scores obtained from the measures described in Objectives One and Two;

(4) To investigate the validity of the newly-developed variables, indicated by their relationship with other appropriate external criteria of creativity

(5) To employ the newly-developed variables in the Creative Studies Project in order to evaluate their unique contributions to the assessment of the effects of the experimental program.

In this Chapter, we shall summarize the findings of the project in relation to those specific objectives.

Objective One

Objective one called for the development of new scoring criteria for the College Situations Problems. Four new scoring criteria were successfully developed for College Situations Problems. The new variables assessed fluency (the number of ideas produced in response to the problem), flexibility (the variety or kinds of ideas produces), originality (the ability to produce unusual or infrequent ideas about the problem) and structural analysis (involving the organization and sequence of ideas produced). Two additional variables were developed, but were subsequently dropped because of problems in attaining stable, objective scores.

Objective Two

Objective two involves the development of a new, realistic but complex measure of creative problem-solving abilities. The Improving Research Testing measure was developed to meet this objective. In this measure, subjects were asked to employ creative problem-solving skills to produce a plan for improving testing procedures for creativity research. The problem was presented in three parts which stressed, respectively, ideation (listing ideas for possible improvements), development of evaluation criteria (for assessing the quality or applicability of the ideas listed in part one), and, in the most complex section, creation of a systematic plan for improving the testing procedures. From these three stages of the problem, five new scores or variables were developed, assessing (1) fluency (number of ideas), (2) fluency of producing criteria for evaluating ideas, (3) workability (potential for actual implementation), (4) importance (value of the proposed plan and changes), and (5) an overall rating of the plan by three expert judges.

Objective Three

Objective three concerned the interrelationships among the newly-developed variables The evidence suggested that the nine new variables could be scored reliably by college students with brief training. In addition, the data indicated that the five variables from the Improving Research Testing measure were positively and significantly intercorrelated. The four variables from the College Situations Problem were also significantly and positively intercorrelated. In general, with only the exception of the fluency measures, the five Improving Research Testing variables

were significantly intercorrelated among themselves, as were all four College Situations Problem variables, while correlations between those two sets of variables were negligible. That is, with the logical exception of fluency scores, variables derived from one measure were more like each other than they were similar to the variables derived from the other measure. We concluded that this was desirable, since we intended that the new measures would yield somewhat different information about creative problem-solving skills and abilities.

Objective Four

Objective four concerned investigations of the validity of the newly-developed variables, indicated by relationships with selected external criteria of creativity. In general, the results indicated that there was some support for the validity of many of the nine new variables. The fluency scores for parts I and II of Improving Research Testing and for the College Situations Problem were positively and significantly correlated with several divergent production measures, and with some areas of non-academic accomplishment (particularly areas involving social activities and humanities). In addition, those subjects who listed the most non-academic accomplishments also tend to have the highest fluency scores on the College Situations Problem.

The Overall Rating for the Improving Research Testing plan was also significantly and positively correlated with several divergent-production criteria, with one transformation criterion -- evaluation of symbolic transformations -- and was significantly negatively correlated with artistic accomplishments. The Originality score for the College Situations Problem was significantly and positively related to many divergent production variables, to one transformation criterion -- memory of semantic transformations -- and to total non-academic accomplishments. The College Situations Problem, Structural Analysis score was significantly and positively related to several non-academic accomplishment areas, including social participation and service, social science, and writing.

The least support was obtained for the Workability and Importance scores for Improving Research Testing, each of which was significantly correlated with only one of the 23 external criteria, and the Flexibility score for the College Situations Problem, which was significantly correlated with only one divergent production measure. In addition, subjects who claimed the fewest (total) non-academic accomplishments attained significantly greater Workability scores than subjects with the highest total number of accomplishments. Although significant negative relationships may have as much predictive power as positive relationships of corresponding magnitude, we were unable to recognize any clear logical explanation for the negative relationship; the result is puzzling and the scores should certainly be regarded very cautiously.

It must also be noted that, although many correlation coefficients were positive and significantly greater than zero, the magnitude of the relationships tended to be "low to moderate." In very few cases did the correlations exceed .30. Thus, we conclude that our results provide preliminary indications supporting the validity of the new variables, although more extensive validation is required, involving additional and more complex external criteria. The divergent production measures, for example, represented only partially the total divergent production category, and did not sample adequately the full range of contents and products. In addition, Guilford (1971) has warned that correlations of complex criteria with only one (or successively with only a few) divergent production categories may yield only moderate correlations. The use of larger test batteries and samples, which would make possible more effective multivariate analyses, must also be planned in validating complex criteria (cf., Treffinger and Poggio, 1972). For the non-academic accomplishment scales, it should also be noted that the scores on each

scale tended to have rather restricted ranges, which may have caused correlations with the new criteria to be somewhat lower than might be expected if a wider range of accomplishment scores had been attained. In general, then, we conclude that the evidence for the validity of the nine new variables is probably somewhat conservative, and is certainly encouraging.

Objective Five

The fifth objective was concerned with the application of the nine new variables in the Creative Studies Project (Parnes and Noller, 1973). For each of the nine variables, then, the four experimental groups and the control group were compared.

Significant differences were obtained for the comparisons employing three of the nine new variables: Improving Research Testing, Fluency, Parts I and II; and the Overall Rating for the Improving Research Testing plan. Among the new variables, these appeared to be stable, and were supported most broadly in the validation investigations. It was determined that the best performance on each of these three measures was attained by the four-semester experimental group, whose scores were significantly greater than the scores of both the controls and the other experimental groups. All experimental groups' mean scores (except for the one-semester group for the Overall Rating variable) were significantly greater than the controls' mean scores on all three of these variables.

Thus, there was evidence for the sensitivity of the new measures in identifying differences among the experimental and control groups in the Creative Studies Project. Despite the complexity of the problems presented, and the development of entirely new scoring procedures, the results of these analyses were remarkably consistent with the general pattern of the results of the project (cf. Parnes and Noller, 1973). These measures provided further support, then, for the value and cumulative effects of instruction in creative problem-solving among college undergraduates.

The direction of the actual scores, even for analyses in which no significant differences were found, appeared also to be consistent with the general findings for the project to date. While tentative, of course, this suggests that continued work on the further development and validation of such criteria, especially including scoring criteria for the College Situations Problems, may be very valuable. In particular, it would probably be quite valuable to utilize the College Situations Problems scoring criteria across a more extensive set of problem situations, since the scores derived for this study (because of time and scoring budget limitations) were restricted to responses to only one problem situation.

Directions for Future Research

Although we have concluded from this project that useful, complex measures can be developed, from which several valid and reliable variables can be derived, several possible dimensions for future research can be identified.

First, multivariate studies of the validity and structure of the new variables should be undertaken. Although such procedures require greater sample sizes than were available for this project (which also increases other practical problems, such as test administration and scoring), their utilization would certainly contribute to our understanding of the validity and reliability of the new variables.

Next, replications would be valuable, in order to determine whether the relationships identified in this study do in fact represent accurate estimates of the population's parameters. Long-term studies of the predictive validity of the new variables, and more extensive concurrent validation against non-cognitive

(personality) criteria would also be valuable.

Finally, further development and research might profitably be addressed to the revision and implementation of the scoring techniques that were dropped from the present study.

Summary and Conclusions

The purposes of this project were to develop new variables for assessing the effectiveness of instruction in creative problem-solving, to investigate the reliability and validity of such variables, and to utilize the variables in the evaluation of the Creative Studies Project.

It is concluded that:

(1) Complex problem situations can yield reliable and valid indices of creative problem-solving abilities;

(2) Application of such indices in the evaluation of the Creative Studies Project supports the effectiveness of the instructional program (especially among subjects who participate for four semesters); and

(3) Additional research, involving multivariate statistical procedures and long-term studies, is necessary in order to provide more comprehensive assessments of the validity of any criteria which purport to measure creative problem-solving abilities.

REFERENCES

Asher, J. W., Feldhusen, J. F., Gruen, G. E., Kane, R. B., McDaniel, E. D., Stevens, M. I., Towler, J., and Wheatley, C. Development of new measures of cognitive variables in elementary school children. (Phase I) Lafayette: Purdue University, 1970. Final report of contract OEC - 0 - 70 - 4952.

Ausubel, P. P. Creativity, general creative abilities and the creative individual. Psychology in the Schools, 1964, 1, 344-347.

Britton, R. J. A study of creativity in selected sixth-grade groups. Unpublished Ed.D. thesis, University of Virginia, 1967.

Clark, P. M. and Mirels, H. L. Fluency as a pervasive element in the measurement of creativity. Journal of Educational Measurement, 1970, 7, 83-86.

Coelho, G. C., Silber, E., and Hamburg, D. A. Use of the student - TAT to assess coping behavior in hospitalized, normal, and exceptionally competent college freshmen. Perceptual and Motor Skills, 1962, 14, 355-365.

Covington, M. V. New directions in the appraisal of creative thinking. In: Treffinger, D. J. (ed.) Readings on creativity in education. Englewood Cliffs: Prentice-Hall, in press.

Covington, M. V., Crutchfield, R. S., and Davies, L. The productive thinking program. Berkeley: Educational Innovations, 1966.

Covington, M. V. and Crutchfield, R. S. Facilitation of creative problem solving. Programmed instruction, 1965, 4, 3-5, 10.

Cunnington, B. F. and Torrance, E. P. Sounds and images. Boston: Ginn, 1965.

Davis, G. A. Current status of research on human problem solving. Psychological Bulletin, 1966, 66, 36-51.

Davis, G. A. and Houtman, S. E. Thinking creatively. Madison, Wisconsin: University of Wisconsin, Research and Development Center for Cognitive Learning, 1968.

Dellas, M. and Gaier, E. L. Identification of creativity: the individual. Psychological Bulletin, 1970, 73, 55-73.

Feldhusen, J. F., Treffinger, D. J., and Thomas, S. J. Global and componential evaluation of creativity instructional materials. Buffalo: Creative Education Foundation, 1971.

Feldhusen, J. F., Treffinger, D. J., and Bahlke, S. J. Developing creative thinking: The Purdue Creativity Training Program. Journal of Creative Behavior, 1970.

Goldfried, M. R. and D'Zurilla, T. J. A behavioral analytic model for assessing competence. In: Spielberger, C. D. (Ed.) Current topics in clinical and community psychology, Volume I. New York: Academic Press, 1969, pp. 151-196.

Gordon, W. J. J. The metaphorical way of knowing. Cambridge, Massachusetts: Porpoise Books, 1971.

Gough, H. G. and Heilbrun, A. B., Jr. The Adjective Check List, Palo Alto: Consulting Psychologist Press, 1965.

Grube, T. F. To make a difference in five days... Journal of Creative Behavior, 1971, 5, 74-75.

Guilford, J. P. Some misconceptions regarding measurement of creative talents. Journal of Creative Behavior, 1971, 5, 77-87.

Guilford, J. P. Creativity: retrospect and prospect. Journal of Creative Behavior, 1970, 4, 149-168.

Guilford, J. P. Creativity: yesterday, today and tomorrow, Journal of Creative Behavior, 1967, 1,.3-14. (b)

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967. (a)

Guilford, J. P. Creativity. American Psychologist, 1950, 5, 444-454.

Guilford, J. P. and Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Guilford, J. P., Merrifield, P. R., Christenson, P. R., and Frick, J. W. The role of intellectual factors in problem solving. Psychological monographs, 1962, 76, 10 (wh. #529)

Harvey, A. J. A partial evaluation of Torrance's Test of Creativity. American educational research Journal, 1970, 7, 359-371.

Khatena, J. Some problems in the measurement of creative behavior, Journal of Research and Development in Education, 1971, 4, 74-82.

MacKinnon, D. W. The nature and nurture of creative talent. American Psychologist, 1962, 17, 484-495.

Meadow, A. and Parnes, S. J. Evaluation of training in creative problem solving. Journal of Applied Psychology, 1959, 43, 189-194.

Mednick, S. A. The associative bases of the creative process. Psychological Review, 1962, 69, 220-232.

Miles, D. T. Development of a test for an experimental research program in creative problem solving. Carbondale, Illinois: Southern Illinois University, Final Report of Project ER-7-E-037, USOE Grant OEG-3-7-700037-2940, 1968.

Myers, R. and Torrance, E. P. Invitations to speaking and writing creatively. Boston: Ginn (Ideal Books Series), 1965.

Olton, R. M. A self-instructional program for developing productive thinking skills in fifth and sixth-grade children. Journal of Creative Behavior, 1969, 3, 16-25.

Osborn, A. F. Applied imagination. New York: Scribner's, 1953.

Parnes, S. J. Creative behavior guidebook. New York: Scribner's, 1967. (a)

Parnes, S. J. Creative potential and the educational experience. Buffalo: Creative Education Foundation, Occasional Paper #2, 1967.(b)

Parnes, S. J. Programming Creative Behavior. Final report of NDEA Title VII Project 5-0716. Buffalo: State University of New York, 1966.

Parnes, S. J. and Meadow, A. Evaluation of persistence of effects produced by a creative problem solving course. Psychological Reports, 1960, 7, 357-361.

Parnes, S. J. and Meadow, A. Effects of "brainstorming" instructions on creative problem solving in trained and untrained subjects. Journal of Educational Psychology, 1959, 50, 171-176.

Parnes, S. J. and Noller, R. B. Toward supersanity: channeled freedom. Buffalo: D.O.K., 1973.

Richards, J. M., Jr., Holland, J. L., and Lutz, S. W. The assessment of student accomplishment in college. ACT Research Reports, No. 11. Iowa City: American College Testing Program, 1966.

Ripple, R. E. and Dacey, J. S. The facilitation of problem solving and verbal creativity by exposure to programmed instruction. Psychology in the schools, 1967, 4, 241-245.

Taft, R. and Rossiter, J. R. The Remote Associates Test: convergent or divergent thinking? Psychological Reports, 1966, 19, 1313-1314.

Thorndike, R. L. Some methodological issues in the study of creativity. In: Proceedings of the 1962 invitational conference on testing problems. Princeton, New Jersey: Educational Testing Service, 1963.

Torrance, E. P. Torrance Tests of Creative Thinking. Princeton: Personnel Press, 1966.

Torrance, E. P. Rewarding creative behavior. Englewood Cliffs: Prentice-Hall, 1965.

Torrance, E. P. Guiding creative talent. Englewood Cliffs: Prentice-Hall, 1962.

Torrance, E. P. and Gupta, R. Development and evaluation of recorded programmed experiences in creative thinking in the fourth grade. Minneapolis: Bureau of Education and Research, U. of. 1964.

Torrance, E. P. and Myers, R. Creative teaching and learning. New York: E. P. Dutton, 1970.

Treffinger, D. J. and Gowan, J. C. An updated representative list of methods and educational programs for stimulating creativity. Journal of Creative Behavior, 1971, 5, 127-139.

Treffinger, D. J. and Poggio, J. P. Needed research on the measurement of creativity. Journal of Creative Behavior, 1972, 6, 253-267.

Treffinger, D. J., Renzulli, J. S. and Feldhusen, J. F. Problems in the assessment of creative thinking. Journal of Creative Behavior, 1971, 5, 104-112.

Treffinger, D. J. and Ripple, R. E. Programmed instruction in creative problem solving. Educational leadership, 1971, 28, 667-675.

Treffinger, D. J. and Ripple, R. E. Developing creative problem solving abilities and related attitudes through programmed instruction. Journal of Creative Behavior, 1969, 3, 105-110, 127.

Tryk, H. E. Assessment in the study of creativity. In: McReynolds, P. (Ed.) Advances in psychological assessment: Volume one. Palo Alto: Science and Behavior Books, 1968, pp. 34-54.

Vernon, P. E. Creativity and intelligence. Educational research, 1964, 6, 163-169.

Wallach, M. A. Review of the Torrance Tests of Creative Thinking. American Educational Research Journal, 1968, 5, 272-281.

Wallach, M. A. and Kogan, N. Modes of thinking in young children. New York: Holt, Rinehart, and Winston, 1965.

Wallach, M. A. and Wing, C. The talented student. New York: Holt, Rinehart, and Winston, 1969.

Wilson, R. C., Guilford, J. P., and Christensen, P. R. The measurement of individual differences in originality. In: Parnes, S. J. and Harding, H. F. (Eds.) A Source book for creative thinking. New York: Scribner's, 1962, 239-247.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

	ORGANIZATION AND STRUCTURE	ANALYSIS AND SYNTHESIS OF IDEAS	UNIQUENESS AND ORIGINALITY	ADAPTIVE TO REALITY	SENSITIVITY AND EMPATHY	APPENDIX A
0	ENTIRELY MISSING (BLANK PAPER): OR IRRELEVANT TO PROBLEM: OR HOSTILE OR DEFENSIVE STATEMENT TO EXAMINER.					
1	Vague general comments - little or no structure. Deals only superficially with problem. Just repeats one or a few simple ideas.	Merely repeats the problem or the situation, with no plan for how the problem could be solved.	Gives only the very obvious answers (i.e., those given by 2/3 or more of the group). Common, banal ideas.	Unreal fantasy - outer space, magic, wishing, psychic or mystical. Completely unusable.	Expresses Disinterest in Problem. Entirely personal "complaints".	Doesn't recognize any new problems which could arise.
2	Response in prose form, but not clearly structured or organized. Simple ideas, strung together.	Repeats the problem, but includes a few ideas for solution (unelaborated and not ordered; just present in a list).	Gives mostly the very obvious answers (2/3 of the group would give). Trivial, unimportant.	Real but totally impractical - "Fatal flaw". Contains no suggestion as to how it might be applied. Wild ideas. Serious defects.	Describe his own feelings about shortcomings and asserts that others feel the same.	
3	Organized response in prose or outline form. Has clear sequence of actions. Does not break problem into major components or sub-problems.	Repeats the problem and sequentially lists several specific actions which could be taken (without much or any elaboration).	Gives ordinary responses - (given by 1/3 but less than 2/3 of the group). Should not be all trivial ideas.	Usually minor or trivial. Possible, but contains some practical or technical flaw (serious, but non-fatal). (Time, cost, bias).	Makes no statement of interest or disinterest. Complaints stated for self and general, but accompanied by specific recommendation for change.	
4	Clear sequence - follows outline or plan. Deals with complex aspects of problem, not just simple. May name sub-problem.	Defines several parts of the problem and presents a specific action for solving each part.	Uncommon answers given by more than 5 others, but less than 1/3 of the group. Definitely includes at least one important aspect of the problem.	Plausible. Flaws are minor. Either: ways of handling problems which arise are given, OR, deals primarily with major issues.	Recognizes positive and negative feelings of self and others. Sees that both + and - can lead to ideas for change. Some ideas are improvements, not just corrections.	
5	Clear, explicit plan - sequence of ideas. Deals with complex issues. States sub-problems and collects ideas for each.	Defines several aspects or sub-problems and proposes more than one possible course of action for each, and shows some integration or synthesis, implicitly or explicitly.	Unusual ideas. (given by 1-5 others in the group at most). Mostly important ideas.	Plausible and workable - could be implemented with few or no changes. Deals with major issues. May name evaluation criteria.	Expresses interest involvement in problem. Sees both + and - ideas as source of change. May see new problems which might arise. Sees the purposes of the experiment.	
6	Very detailed plan includes generalization, sub-problems, and well-developed ideas for solving each sub-problem. Tightly organized.	Defines several aspects or sub-problems. Presents several possible actions for each. Synthesizes them into a more comprehensive plan. May redefine problem.	Extremely Unique Ideas--given by one subject only. (Response may also contain other ideas).	Well-defined ways of applying the ideas in a real situation. Suggests multiple alternatives.	(All of #5) plus discusses ways of dealing with feelings. Sees problems and solutions.	

Appendix B

Means and Standard Deviations for all Project Variables (N=158)

Variable	$\bar{X}$	S.D.
Structure-of-Intellect Tests		
Alternate Letter Groups	6.53	2.08
Multiple Social Problems	9.42	2.70
Insight Problems	1.64	1.06
Utility: Fluency	17.88	4.85
Utility: Flexibility	12.30	4.38
Verbal-Pict. Translation	9.38	2.65
Homonyms	7.21	2.29
Jumbled Words	23.84	2.69
Multiple Behavioral Group	2.60	1.16
Varied Emotional Rel.	9.61	2.22
Selected ACL Scale	11.82	5.44
Non-Academic Accomplishment Scales		
Leadership	0.99	1.54
Social Participation	1.14	1.65
Artistic	1.65	1.77
Social Service	1.46	1.77
Scientific	0.15	0.54
Business	0.70	1.01
Humanistic/Cultural	1.88	1.58
Religious	1.36	2.12
Musical	0.29	0.71
Writing	0.96	1.03
Social Service	0.50	0.81
Speech and Drama	0.55	1.12
Total No. of Items	11.75	7.98
Improving Research Testing		
Fluency, Part I	7.00	4.45
Fluency, Part II	5.81	3.22
Workability	3.57	2.66
Importance	1.98	0.66
Overall Rating (x 10)	27.05	14.19
College Situations Problems		
Fluency	3.53	1.86
Flexibility	2.67	1.11
Originality	4.82	3.21
Structural Analysis	1.89	0.65